Mario Baseball Data Analysis

Iszy Hirschtritt Licht

12/1/2020

Load Libraries

```
library(tidyverse)
library(dplyr)
library(knitr)
library(weights)
library(scales)
library(stargazer)
library(ggthemes)
```

Load Data

```
#Load Data
mario_data <- read.csv("Mario_Baseball_Data.csv")

#Clean Data
mario_data <- mario_data %>%
  replace(is.na(.), 0) %>%
  rename(
    date = Date,
    player_name = Player.Name,
    played_game = Games.Played,
    at_bats = AB,
    hits = Hits,
    runs_batted_in = RBI,
    homeruns = HR,
    stolen_bases = SB,
    special_hitting = Special,
    innings_pitched = IP,
    hits_allowed = Hits.1,
    runs_allowed = Runs,
    strikeouts = SO,
    big_plays = Big.Plays,
    special_pitching = Special.1,
    player_type = Player.Type,
    captain = Capitan
    ) %>%
  mutate(date = as.Date(date,"%m.%d.%y")) %>%
  mutate(played_game = as.factor(played_game)) %>%
  mutate(captain = as.factor(captain))
```

Data Analysis

```
#Add Rate Data to Dataset
mario_data <- mario_data %>%
  group_by(player_name) %>%
  mutate(
    special_use_rate = sum(special_hitting)/sum(at_bats),
    batting_average = sum(hits)/sum(at_bats),
    era = (sum(runs_allowed)/sum(innings_pitched)*9),
    so9 = (sum(strikeouts)/sum(innings_pitched)*9),
    hip = sum(hits_allowed)/sum(innings_pitched))

#By Player Hitting
player_hitting <- mario_data %>%
  group_by(player_name) %>%
  summarise(
    batting_average = sum(hits)/sum(at_bats),
    special_use_rate = sum(special_hitting)/sum(at_bats)
  )
kable(player_hitting, digits = 3)
```

player_name	batting_average	special_use_rate
Baby Bowser	0.147	0.059
Baby Luigi	0.176	0.059
Baby Mario	0.276	0.000
Birdo	0.343	0.260
Boo	0.343	0.000
Bowser	0.332	0.035
Daisy	0.345	0.152
Diddy Kong	0.053	0.105
DK	0.409	0.100
Drybones	0.286	0.006
Flying Goomba	0.000	0.000
Flying Koopa	0.325	0.007
Goomba	0.316	0.000
Grandpapa Toad	0.438	0.000
Hammer/Etc. Bro	0.380	0.000
King Boo	0.209	0.015
Koopa	0.313	0.010
Luigi	0.336	0.043
Magikoopa	0.218	0.007
Mario	0.454	0.430
Monty	0.194	0.000
Mumbo	0.248	0.000
Noki	0.234	0.065
Peach	0.254	0.099
Petey	0.295	0.000
Shy Guy	0.171	0.000
Toad	0.391	0.000
Toadette	0.211	0.000
Waluigi	0.333	0.190
Wario	0.182	0.091
Yoshi	0.346	0.132

```r
#By Player Type Hitting
player_type_hitting <- mario_data %>%
  group_by(player_type) %>%
  summarise(
    total_ab = sum(at_bats),
    total_hits = sum(hits),
    total_runs_batted_in = sum(runs_batted_in),
    total_hits = sum(hits),
    total_homeruns = sum(homeruns),
    total_sb = sum(stolen_bases),
    batting_average = sum(hits)/sum(at_bats),
    special_use_rate = sum(special_hitting)/sum(at_bats),
    sb_hits = total_sb/total_hits
    )
kable(player_type_hitting, digits = 3)
```

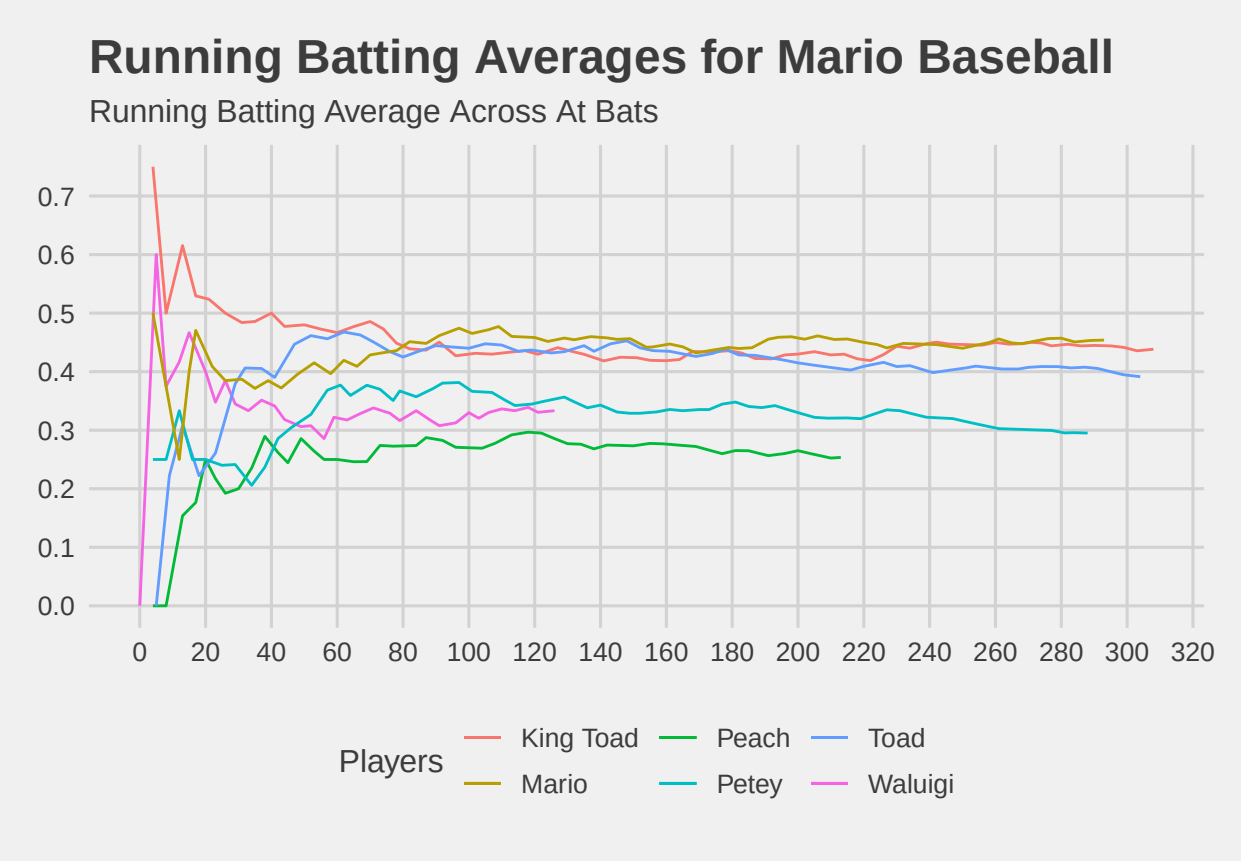
player_type	total_ab	total_hits	total_runs_batted_in	total_homeruns	total_sb	batting_average	special_us
Balance	1798	646	195	6	61	0.359	
Power	1368	448	181	43	29	0.327	
Speed	485	136	36	2	17	0.280	
Technique	1537	504	161	4	41	0.328	

```r
#By Y/N Captain Hitting
captain_stats <- mario_data %>%
  group_by(captain) %>%
  summarise(batting_average = sum(hits)/sum(at_bats))
kable(captain_stats, digits = 3)
```

captain	batting_average
0	0.335
1	0.300

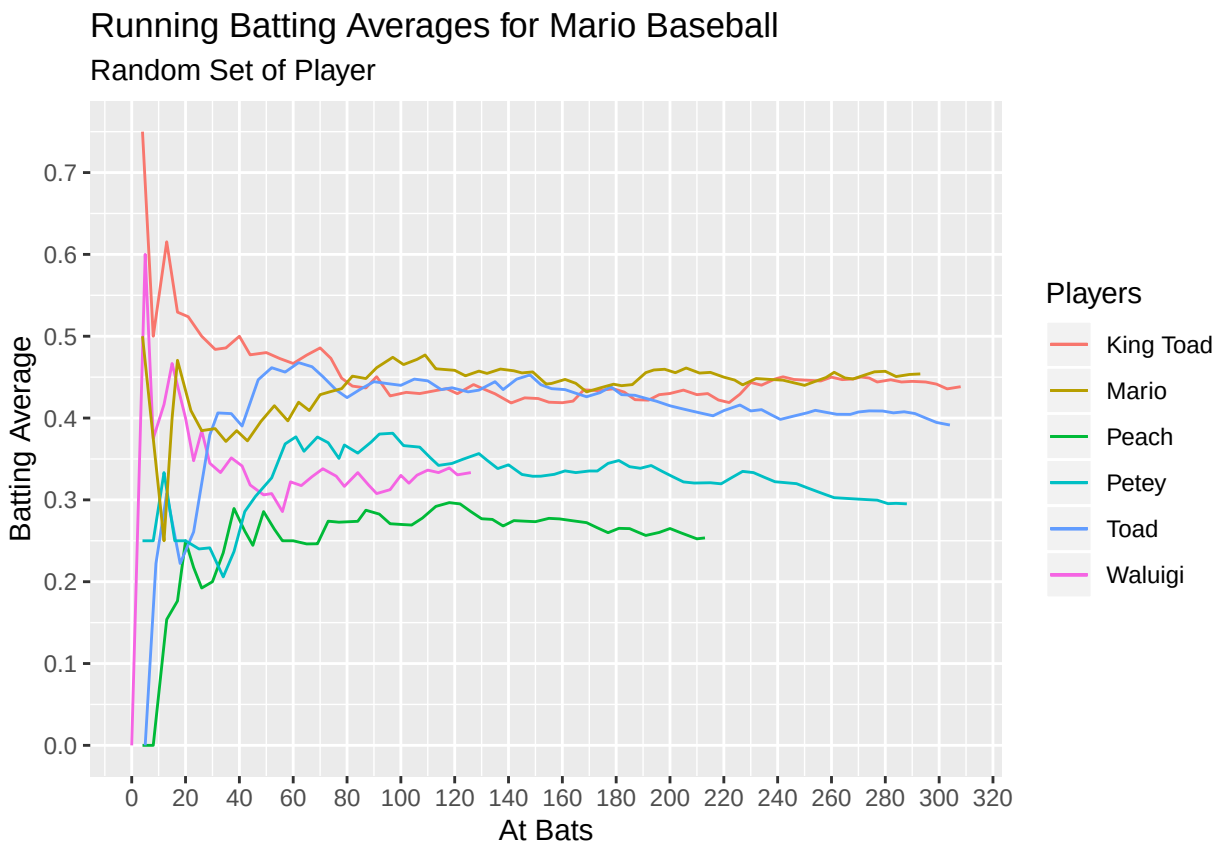
```r
#Running Batting Averages
mario_data <- mario_data %>%
  mutate(
    cum_at_bats = cumsum(at_bats),
    cum_hits = cumsum(hits),
    running_avg = cum_hits / cum_at_bats) %>%
  replace(is.na(.), 0)

#Plot Running Batting Averages
king_toad <- mario_data %>%
  filter(player_name == "Grandpapa Toad")

waluigi <- mario_data %>%
  filter(player_name == "Waluigi")

peach <- mario_data %>%
  filter(player_name == "Peach")
```

```r
toad <- mario_data %>%
  filter(player_name == "Toad")

petey <-  mario_data %>%
  filter(player_name == "Petey")

mario <-  mario_data %>%
  filter(player_name == "Mario")

#Plot of 6 Players
ggplot() +
  geom_line(king_toad, mapping = aes(x = cum_at_bats, y = running_avg, color = "King Toad")) +
  geom_line(waluigi, mapping = aes(x = cum_at_bats, y = running_avg, color = "Waluigi")) +
  geom_line(peach, mapping = aes(x = cum_at_bats, y = running_avg, color = "Peach")) +
  geom_line(toad, mapping = aes(x = cum_at_bats, y = running_avg, color = "Toad")) +
  geom_line(petey, mapping = aes(x = cum_at_bats, y = running_avg, color = "Petey")) +
  geom_line(mario, mapping = aes(x = cum_at_bats, y = running_avg, color = "Mario")) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 20)) +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 10)) +
  labs(title = "Running Batting Averages for Mario Baseball",
       subtitle = "Running Batting Average Across At Bats", x = "At Bats", y ="Batting Average") +
  scale_colour_discrete("Players") +
  theme_fivethirtyeight()
```



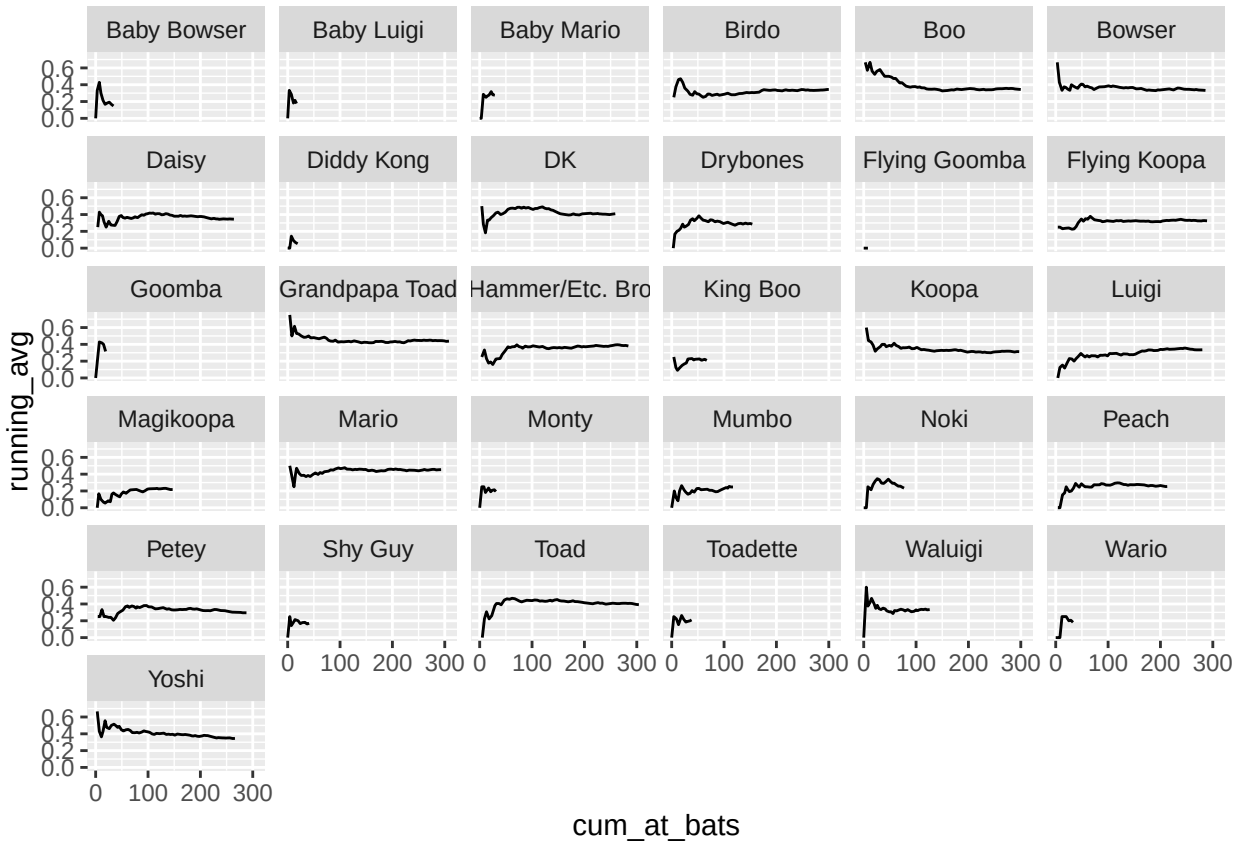

```r
ggplot() +
  geom_line(king_toad, mapping = aes(x = cum_at_bats, y = running_avg, color = "King Toad")) +
```

```
  geom_line(waluigi, mapping = aes(x = cum_at_bats, y = running_avg, color = "Waluigi")) +
  geom_line(peach, mapping = aes(x = cum_at_bats, y = running_avg, color = "Peach")) +
  geom_line(toad, mapping = aes(x = cum_at_bats, y = running_avg, color = "Toad")) +
  geom_line(petey, mapping = aes(x = cum_at_bats, y = running_avg, color = "Petey")) +
  geom_line(mario, mapping = aes(x = cum_at_bats, y = running_avg, color = "Mario")) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 20)) +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 10)) +
  labs(title = "Running Batting Averages for Mario Baseball",
       subtitle = "Random Set of Player", x = "At Bats", y ="Batting Average") +
  scale_color_discrete("Players")
```

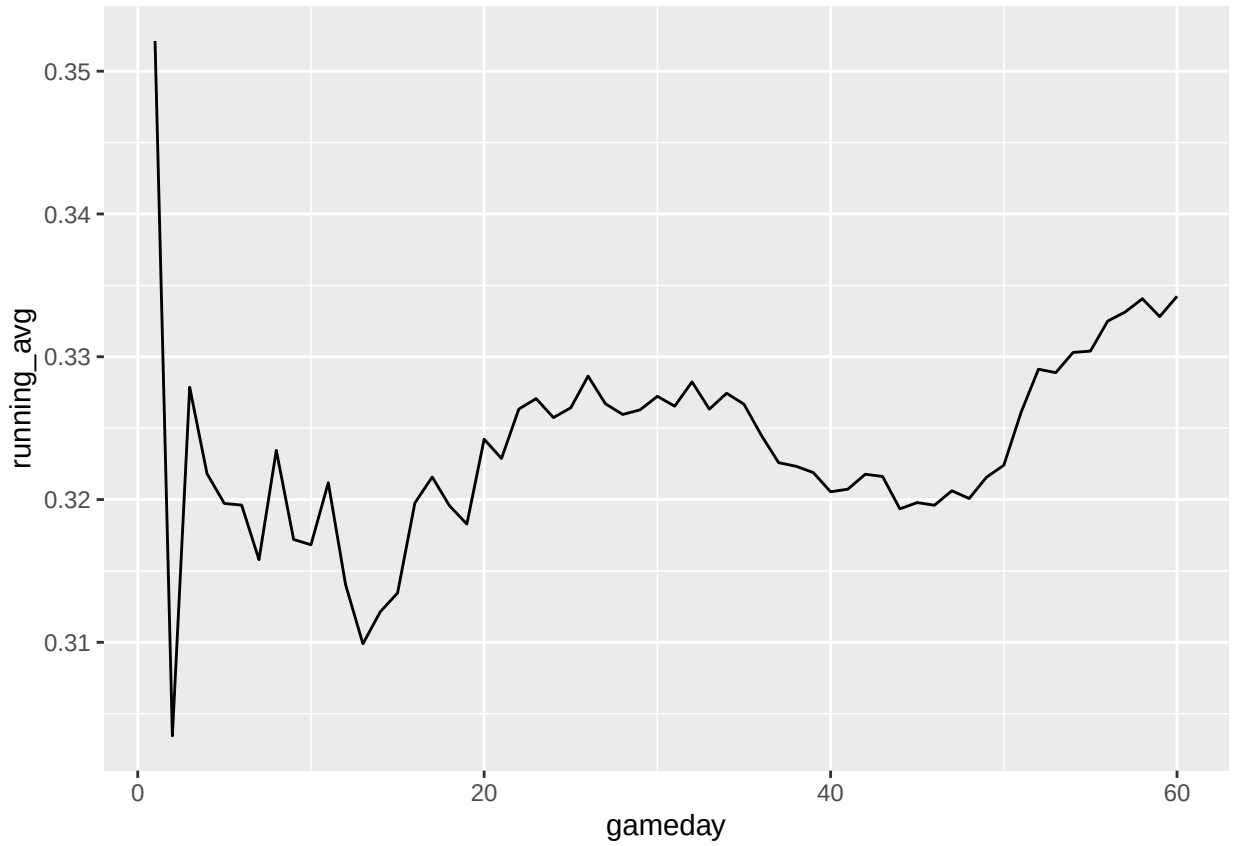



```
# Plot of all Players
ggplot(mario_data, aes(x=cum_at_bats, y=running_avg, group=player_name, shape=player_name)) +
  geom_line() +
  facet_wrap(~ player_name)
```

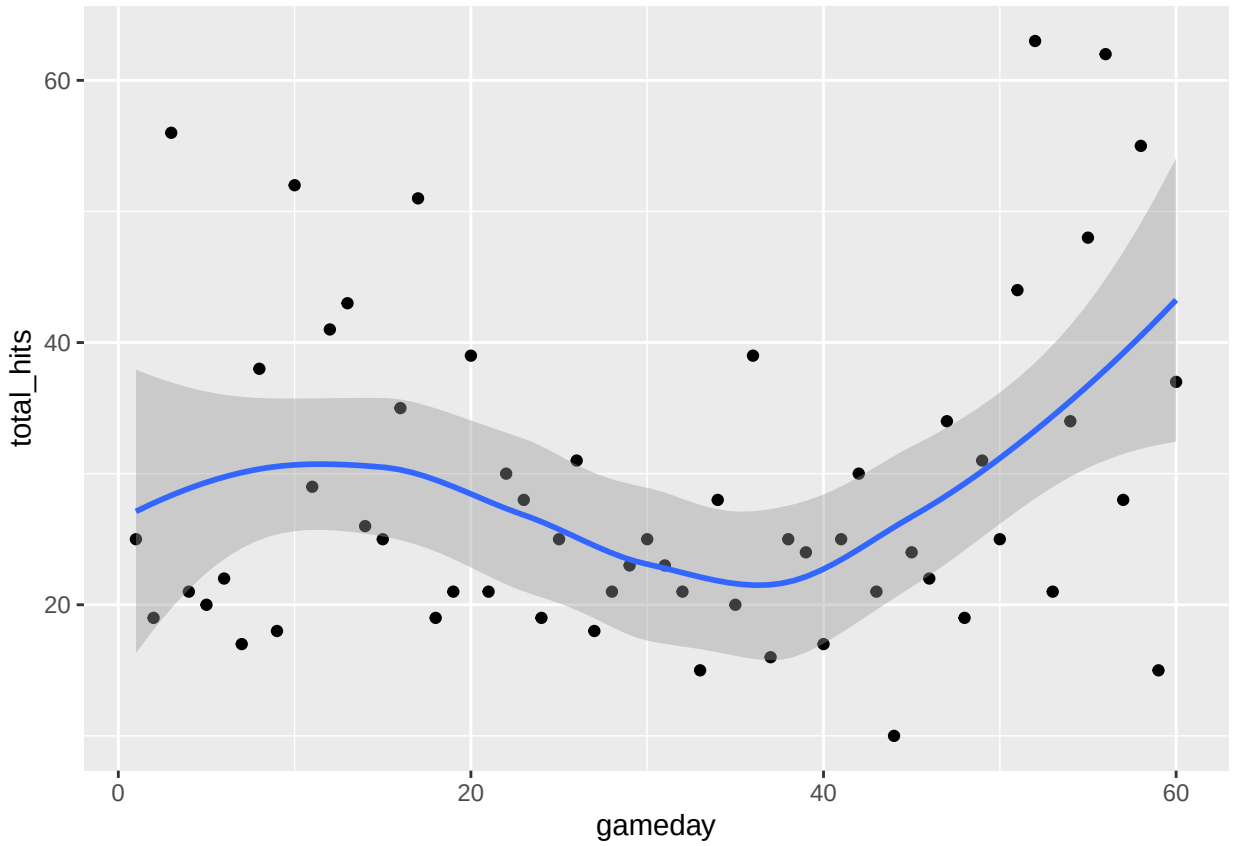
```
## Warning: The shape palette can deal with a maximum of 6 discrete values because
## more than 6 becomes difficult to discriminate; you have 31. Consider
## specifying shapes manually if you must have them.
```




```
#Leagewide running average
leaguewide_data <- mario_data %>%
  group_by(date) %>%
  summarise(
    total_hits = sum(hits),
    total_at_bats = sum(at_bats),
    total_average = sum(total_hits)/sum(total_at_bats)
    ) %>%
  mutate(
    gameday = row_number(),
    cum_at_bats = cumsum(total_at_bats),
    cum_hits = cumsum(total_hits),
    running_avg = cum_hits / cum_at_bats
  )


#Leaguewide Average Plot
ggplot() +
  geom_line(leaguewide_data, mapping = aes(x=gameday, y=running_avg))
```




```
#Leaguewide Hits plot
ggplot(leaguewide_data, mapping = aes(x=gameday, y=total_hits)) +
  geom_point() +
  geom_smooth(method = "loess")
```




```
batting_avg_captain <- lm(batting_average ~ player_type + special_use_rate + captain ,
                      data = mario_data)

stargazer(batting_avg_captain,
          type = "latex", header = FALSE,
          title = "Regression of Batting Average on Player Type with Controls",
          intercept.bottom = FALSE, single.row=TRUE)
```

Table 4: Regression of Batting Average on Player Type with Controls

	Dependent variable:
	batting_average
Constant	0.306^{***} (0.004)
player_typePower	-0.040^{***} (0.005)
player_typeSpeed	-0.131^{***} (0.005)
player_typeTechnique	-0.003 (0.005)
special_use_rate	0.253^{***} (0.020)
captain1	-0.011 (0.017)
Observations	2,170
R^2	0.354
Adjusted R^2	0.353
Residual Std. Error	0.083 (df = 2164)
F Statistic	237.401^{***} (df = 5; 2164)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
#Player Pitching
player_era_1 <- mario_data %>%
  filter(sum(innings_pitched) >= 40) %>%
  group_by(player_name) %>%
  summarise(era = (sum(runs_allowed)/sum(innings_pitched)*9))
kable(player_era_1, digits = 3)
```

player_name	era
Boo	4.388
Flying Koopa	4.130
Koopa	3.207
Waluigi	4.750

```r
player_era_2 <- mario_data %>%
  filter(sum(innings_pitched) >= 10 & sum(innings_pitched) < 40) %>%
  group_by(player_name) %>%
  summarise(era = (sum(runs_allowed)/sum(innings_pitched)*9))
kable(player_era_2, digits = 3)
```

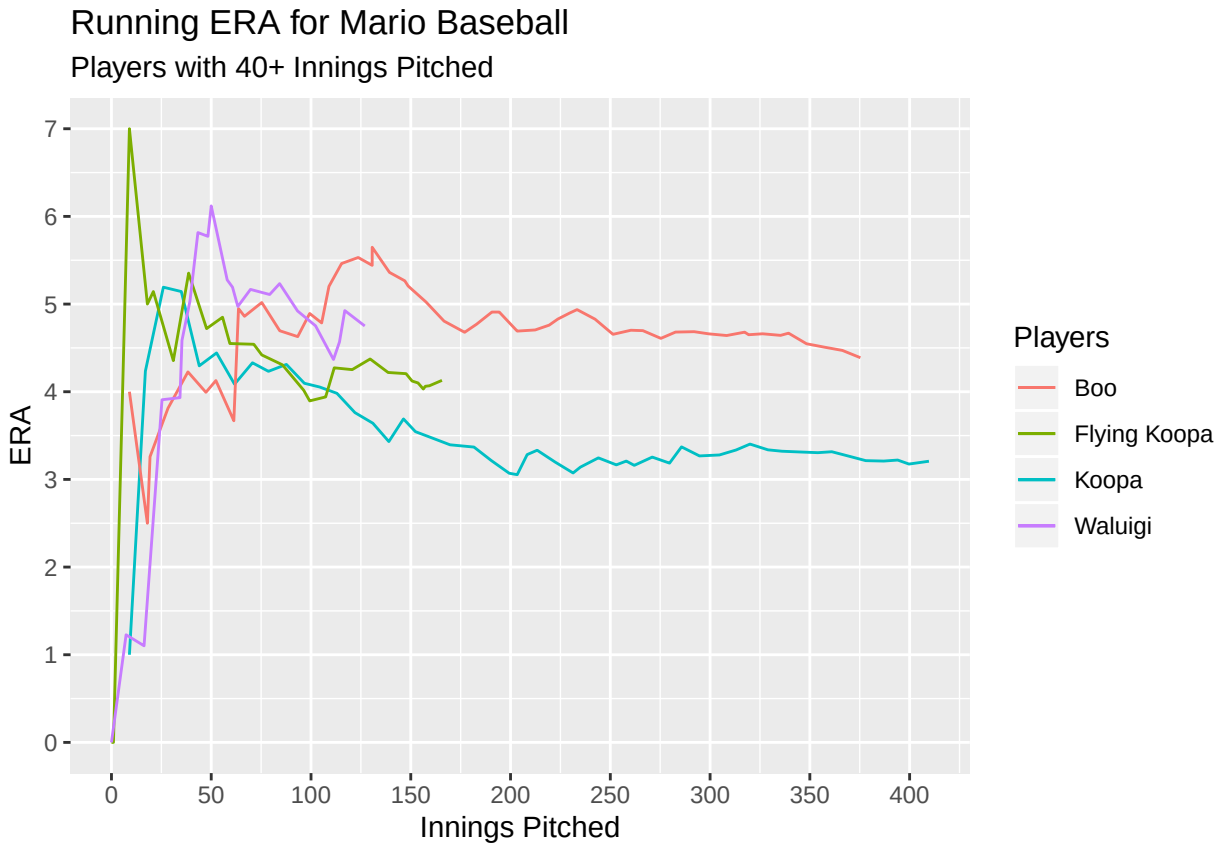
player_name	era
Baby Luigi	3.378
Daisy	4.670
Diddy Kong	2.544
DK	7.200
Peach	6.752
Toad	4.627

```r
#Player Type Pitching
player_type_pitching <- mario_data %>%
  group_by(player_type) %>%
  summarise(
    total_innings = sum(innings_pitched),
    era = (sum(runs_allowed)/sum(innings_pitched)*9),
    total_strikeouts = sum(strikeouts),
    total_big_plays = sum(big_plays),
    )
kable(player_type_pitching, digits = 3)
```

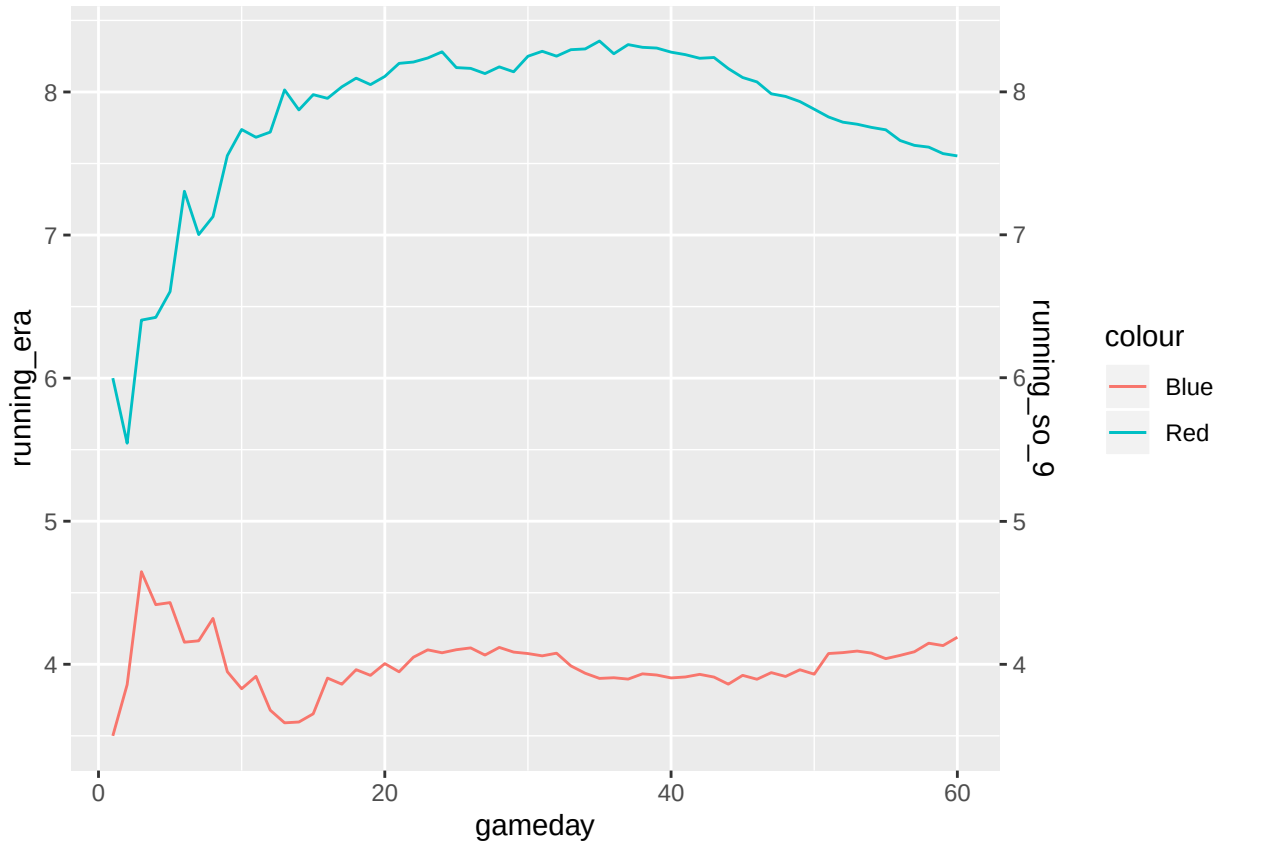
player_type	total_innings	era	total_strikeouts	total_big_plays
Balance	460.597	3.556	407	146
Power	34.650	9.091	25	81
Speed	71.600	3.142	42	41
Technique	687.890	4.475	579	62

```r
#Running Pitching Stats
mario_data <- mario_data %>%
  mutate(
    cum_runs_allowed = cumsum(runs_allowed),
    cum_innings = cumsum(innings_pitched),
```

```
    running_era = (cum_runs_allowed / cum_innings)*9) %>%
  replace(is.na(.), 0)

waluigi <- mario_data %>%
  filter(player_name == "Waluigi")

flying_koopa <- mario_data %>%
  filter(player_name == "Flying Koopa")

koopa <- mario_data %>%
  filter(player_name == "Koopa")

boo <- mario_data %>%
  filter(player_name == "Boo")

#Plot Running Pitching
ggplot() +
  geom_line(koopa, mapping = aes(x = cum_innings, y = running_era, color = "Koopa")) +
  geom_line(boo, mapping = aes(x = cum_innings, y = running_era, color = "Boo")) +
  geom_line(flying_koopa, mapping = aes(x = cum_innings, y = running_era, color = "Flying Koopa")) +
  geom_line(waluigi, mapping = aes(x = cum_innings, y = running_era, color = "Waluigi")) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 10)) +
  labs(title = "Running ERA for Mario Baseball",
       subtitle = "Players with 40+ Innings Pitched", x = "Innings Pitched", y ="ERA") +
  scale_colour_discrete("Players")
```



```r
#Leaguewide Pitching
leaguewide_pitching <- mario_data %>%
  group_by(date) %>%
  summarise(
    total_innings = sum(innings_pitched),
    total_hits_allowed = sum(hits_allowed),
    total_runs_allowed = sum(runs_allowed),
    total_strikeouts = sum(strikeouts),
    total_era = ((sum(runs_allowed)/sum(innings_pitched))*9)
    ) %>%
  mutate(
    gameday = row_number(),
    cum_innings = cumsum(total_innings),
    cum_runs_allowed = cumsum(total_runs_allowed),
    cum_strikeouts = cumsum(total_strikeouts),
    running_so_9 = ((cum_strikeouts/cum_innings)*9),
    running_era = ((cum_runs_allowed / cum_innings)*9))


#Leaguewide ERA Plot
ggplot() +
  geom_line(leaguewide_pitching, mapping = aes(x=gameday, y=running_era, color = "Blue")) +
  geom_line(leaguewide_pitching, mapping = aes(x=gameday, y=running_so_9, color = "Red"))+
   scale_y_continuous(
    "running_era",
    sec.axis = sec_axis(~ . * 1, name = "running_so_9")
  )
```




```
write.csv(mario_data,'Mario_Baseball_Data_update.csv')
write.csv(leaguewide_data,'leaguewide_data.csv')
```